MARGINAL DISTRIBUTION WITH GIBBS SAMPLING ALGORITHM

Anaranya Basu
Instructor: Prof. Manisha Pal

November 2023

1 PROBLEM SET 3

Question: Suppose the joint distribution of Y and θ is given by-

$$P(Y,\theta) = \binom{n}{Y}\theta^{(Y+a-1)}(1-\theta)^{(n-Y+b-1)}, \quad Y = 0,1,2,\ldots,n, \quad 0 < \theta < 1$$

Use Gibbs sampling to find the marginal distribution of Y given n = 16, a = 2 and b = 4. The initial value of θ can be chosen from a U (0,1) distribution.

Solution:

Given,

$$f_Y(k;n,\theta) = \binom{n}{k} \cdot \theta^k \cdot (1-\theta)^{n-k}$$

and, $\pi(\theta) \propto \theta^{x+a-1}(1-\theta)^{n-x+b-1}$
The Joint density of (Y,θ) is -

$$f(Y,\theta)\pi(\theta) = \binom{n}{Y}\theta^{x+a-1}(1-\theta)^{n-x+a-1}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

The Marginal of Y is-

$$f(Y) = \binom{n}{Y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+Y)\Gamma(b+n-Y)}{\Gamma(a+b+n)}$$

The Gibbs Sampling Algorithm is described as follows-
Suppose we have a joint probability distribution $P(X_1, X_2, \ldots, X_n)$ over n variables $X_1, X_2, \ldots, X_n$.

Initialization: Start with initial values for each variable: $(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$.

Iterative Sampling: For $t = 1$ to some predefined number of iterations or until convergence: For each variable X_i:

- Sample $x_i^{(t)}$ from its conditional distribution given the current values of the other variables:

$$x_i^{(t)} \sim P(X_i \mid x_1^{(t)}, x_2^{(t)}, \ldots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \ldots, x_n^{(t-1)})$$

- Update the value of X_i to $x_i^{(t)}$.

Convergence Criteria: Monitor convergence using a suitable criterion, such as assessing the change in sampled values between iterations or using statistical tests.

Termination: Stop the iterations either when convergence is achieved or after a predetermined number of iterations.(Here we have set number of iteration to 1000)

Step 1 :

Now as the conditional distribution of $Y|\theta$ ***is-***

$$f(Y|\theta) = \frac{\binom{n}{Y}\theta^{Y+a-1}(1-\theta)^{n-Y+b-1}}{\sum_{k=0}^{n}\binom{n}{k}\theta^{k+a-1}(1-\theta)^{n-k+b-1}}$$

Step 2:

Now we will Set the initial values of Y *and* θ. *In this case, we can choose the initial value of* θ *from a uniform distribution between 0 and 1 and the initial value of* Y *from a binomial distribution with size* n *and probability* θ.

Step 3:

In Next step- For each iteration of the Gibbs sampler, do the following:

- Sample Y from the conditional distribution of Y given θ.
- Sample θ from the conditional distribution of θ given Y.

In the final step we will Calculate the marginal distribution of Y.

After the Gibbs sampler has run for a sufficiently large number of iterations, the marginal distribution of Y can be approximated by the empirical distribution of the sampled Y values.

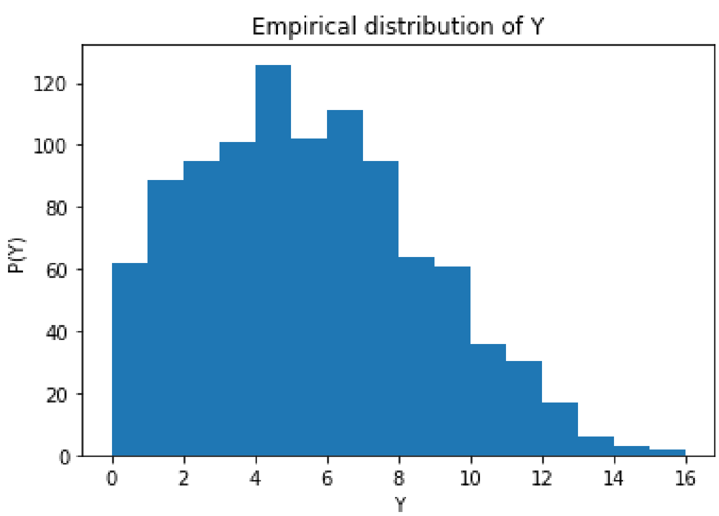
Using R command we get the Gibbs sampler values as follows-

```
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
 34  54  84 119 111  90 114  95  92  69  53  33  20  16   9   4   2
```

And the marginal distribution of Y, which is the probability of each value of Y occurring is-

y	Probability
0	0.020580
1	0.082564
2	0.203356
3	0.274624
4	0.224809
5	0.125616
6	0.053642
7	0.010517
8	0.001329
9	0.000284
10	0.000078
11	0.000017
12	0.000004
13	0.000001
14	0.000000
15	0.000000
16	0.000000

The emperical distribution plot of the marginal of Y is as follows-



The distribution appears to be positively skewed, with a higher concentration of values on the lower end.

The R code is -

```
  n <- 16
a <- 2
b <- 4
# Set the number of iterations
M <- 1000
# Initialize the vectors for storing the samples
y_samples <- rep(NA, M)
theta_samples <- rep(NA, M)
# Initialize the starting values
theta <- runif(1)
y <- rbinom(1, size = n, prob = theta)
# Perform Gibbs sampling iterations
for (i in 2:M) {
  # Sample Y from the conditional distribution P(Y | theta)
  y <- rbinom(1, size = n, prob = theta)

  # Sample theta from the conditional distribution P(theta | Y)
  theta <- rbeta(1, shape1 = a + y, shape2 = b + n - y)

  # Store the samples
  y_samples[i] <- y
  theta_samples[i] <- theta
}

# Calculate the marginal distribution of Y
y_counts <- table(y_samples)
y_probs <- y_counts / M

# Print the marginal distribution of Y
print(y_probs)
```